ExecutionSession3

Deepti Gupta

August 3, 2019

INTRODUCTION

Analyse if there exists a gender gap between male and female incomes between different categories of industry from the labour statistics data from the Bureau of Labor Statistics

DATA ANALYSIS

The x value is equal 5.

```r
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.5.3
```

```r
income <- read_csv("C:/Study/695-ResearchMethodologyAndWriting/Week13/income.csv")
```

```
## Parsed with column specification:
## cols(
##   Occupation = col_character(),
##   Industry = col_character(),
##   All_workers = col_double(),
##   All_weekly = col_double(),
##   M_workers = col_double(),
##   M_weekly = col_double(),
##   F_workers = col_double(),
##   F_weekly = col_double()
## )
```

```r
View(income)

income1 <- income[complete.cases(income),]
head(income1)
```

```
## # A tibble: 6 x 8
##   Occupation Industry All_workers All_weekly M_workers M_weekly F_workers
##   <chr>      <chr>          <dbl>      <dbl>     <dbl>    <dbl>     <dbl>
## 1 Chief exe~ Managem~        1046       2041       763     2251       283
## 2 General a~ Managem~         823       1260       621     1347       202
## 3 Marketing~ Managem~         948       1462       570     1603       378
## 4 Administr~ Managem~         170       1191        96     1451        73
## 5 Computer ~ Managem~         636       1728       466     1817       169
## 6 Financial~ Managem~        1124       1408       551     1732       573
## # ... with 1 more variable: F_weekly <dbl>
```

```r
incomeMale <- income1[,c(2,6)]
incomeMale$Gender <- "Male"
colnames(incomeMale) <- c("Industry", "Income", "Gender")
head(incomeMale)
```

```
## # A tibble: 6 x 3
```

```
##    Industry    Income Gender
##    <chr>         <dbl> <chr>
## 1 Management     2251 Male
## 2 Management     1347 Male
## 3 Management     1603 Male
## 4 Management     1451 Male
## 5 Management     1817 Male
## 6 Management     1732 Male
```

```r
incomeFemale <- income1[,c(2,7)]
incomeFemale$Gender <- "Female"
colnames(incomeFemale) <- c("Industry", "Income", "Gender")
head(incomeFemale)
```

```
## # A tibble: 6 x 3
##    Industry    Income Gender
##    <chr>         <dbl> <chr>
## 1 Management      283 Female
## 2 Management      202 Female
## 3 Management      378 Female
## 4 Management       73 Female
## 5 Management      169 Female
## 6 Management      573 Female
```

```r
incomeAll <- rbind(incomeMale, incomeFemale)
head(incomeAll)
```

```
## # A tibble: 6 x 3
##    Industry    Income Gender
##    <chr>         <dbl> <chr>
## 1 Management     2251 Male
## 2 Management     1347 Male
## 3 Management     1603 Male
## 4 Management     1451 Male
## 5 Management     1817 Male
## 6 Management     1732 Male
```

```r
incomeAll$Industry <- as.factor(incomeAll$Industry)
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```
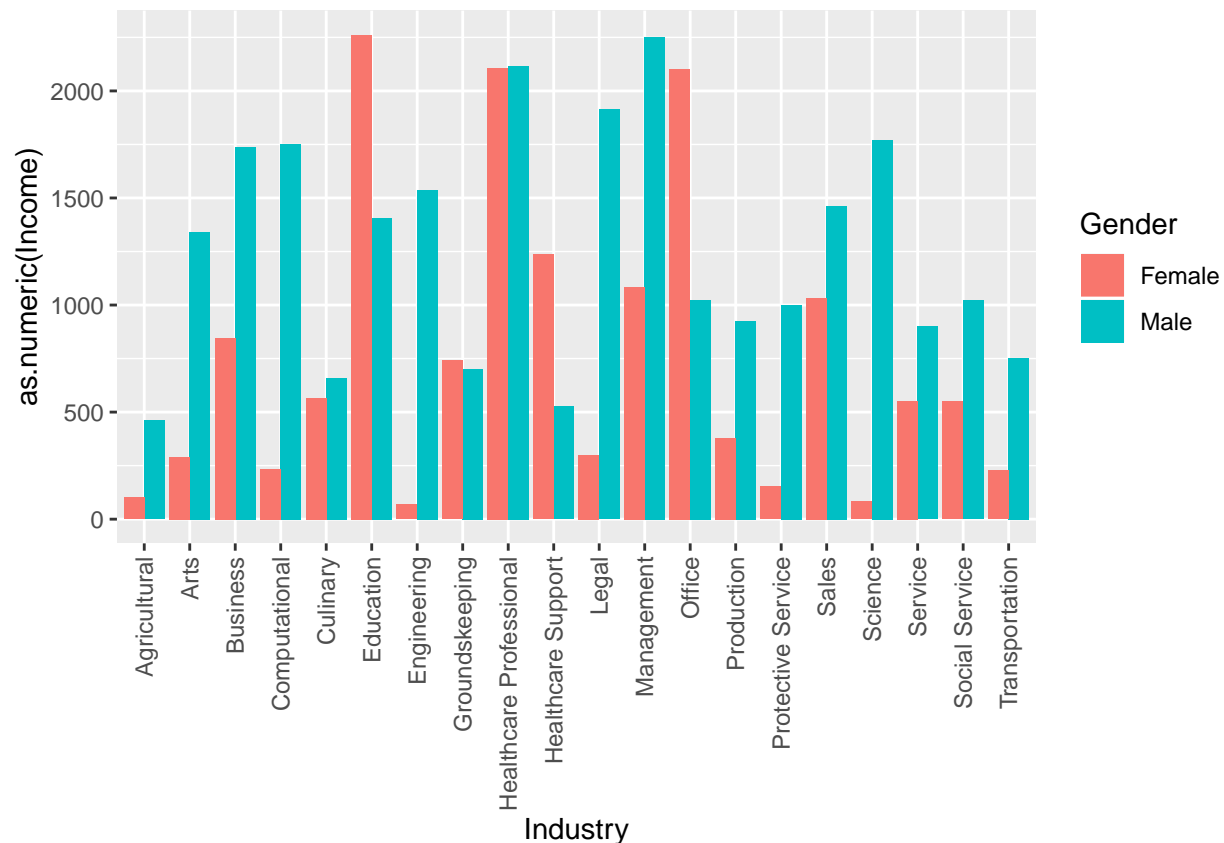
```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.5.3
```

```
## Loading required package: magrittr
```

```
## Warning: package 'magrittr' was built under R version 3.5.3
```

```r
ggplot(data=incomeAll, aes(y=as.numeric(Income), x=Industry, Fill = as.factor(Gender))) + geom_bar(stat=
```

REGRESSION ANALYSIS

```r
lmIncome <- lm(Income ~ ., data = incomeAll)
summary(lmIncome)
```

```
##
## Call:
## lm(formula = Income ~ ., data = incomeAll)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -711.58 -257.93  -70.25  159.39 1847.31
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -75.08     265.83  -0.282   0.7779
## IndustryArts                    392.00     305.67   1.282   0.2011
## IndustryBusiness                493.88     275.53   1.792   0.0744 .
## IndustryComputational           486.36     283.00   1.719   0.0871 .
## IndustryCulinary                102.14     283.00   0.361   0.7185
## IndustryEducation               734.88     295.96   2.483   0.0138 *
## IndustryEngineering             379.75     324.21   1.171   0.2428
## IndustryGroundskeeping          211.17     305.67   0.691   0.4904
## IndustryHealthcare Professional 635.64     283.00   2.246   0.0257 *
## IndustryHealthcare Support      600.50     374.37   1.604   0.1102
```

```
## IndustryLegal                        826.00      374.37    2.206    0.0284 *
## IndustryManagement                    587.72      272.86    2.154    0.0324 *
## IndustryOffice                        326.76      272.39    1.200    0.2316
## IndustryProduction                    116.14      276.49    0.420    0.6749
## IndustryProtective Service            172.83      305.67    0.565    0.5724
## IndustrySales                         413.41      276.49    1.495    0.1363
## IndustryScience                       446.17      305.67    1.460    0.1458
## IndustryService                       191.00      305.67    0.625    0.5327
## IndustrySocial Service                374.33      305.67    1.225    0.2220
## IndustryTransportation                110.25      295.96    0.373    0.7099
## GenderMale                            712.15       48.53   14.673    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 374.4 on 217 degrees of freedom
## Multiple R-squared:  0.56,  Adjusted R-squared:  0.5194
## F-statistic: 13.81 on 20 and 217 DF,  p-value: < 2.2e-16
```